DOCUMENT RESUME

ED 251 501 TM 840 798

AUTHOR Humes, Ann
TITLE What's Involved in Collaborative R & D with a School District.
INSTITUTION Southwest Regional Laboratory for Educational Research and Development, Los Alamitos, Calif.
REPORT NO SWRL-TN-2-83/24
PUB DATE 15 Dec 83
NOTE 33p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS High Schools; *Institutional Cooperation; Interrater Reliability; Minimum Competency Testing; *Regional Laboratories; *School Districts; Scoring; *Test Construction; Test Validity; *Writing Evaluation
IDENTIFIERS *Southwest Regional Laboratory

ABSTRACT

This paper, as an illustration of the procedures involved in a cooperative effort, describes a project in which the Southwest Regional Laboratory (SWRL) designed and developed a minimum standards test in collaboration with a large urban school district in California. The activity described focuses on the writing sample included in the test. The original prompt asked students to write something that they remembered about their impressions on their first day at school. A review revealed that the prompt could pose problems for students who did not remember their first day at school, and that students would likely interpret the purpose of the task differently. SWRL suggested that a new prompt be devised and be tried with a sample of district students. The district selected the sample and the raters. SWRL developed the assessment and scoring materials. The new prompt asked students to write a letter to the fictitious manager of a nearby restaurant, to convince the manager to hire them for a specific job. Several scoring methods were evaluated; SWRL advised the district to use its usual holistic scoring techniques. The feasibility inquiry suggested that the writing sample, as revised, was an appropriate task for the students. The appendix deals with scoring procedures. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

BEST COPY AVAILABLE

ED251501

SWRL

SOUTHWEST REGIONAL LABORATORY
TECHNICAL NOTE

DATE: December 15, 1983

NO: TN 2-83/24

WHAT'S INVOLVED IN COLLABORATIVE R&D WITH A SCHOOL DISTRICT

Ann Humes

ABSTRACT

The characteristics of collaborative R&D with school districts are discussed in general terms. To explain these endeavors, a case example of one project is included: a writing assessment project.

TM 840 798

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P.A. Mills

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

This document is intended for internal staff distribution and use. Permission to reprint or quote from this working document, wholly or in part, should be obtained from SWRL, 4665 Lampson Avenue, Los Alamitos, California 90720.

WHAT'S INVOLVED IN COLLABORATIVE R&D WITH A SCHOOL DISTRICT

Ann Humes

SWRL has a history of successful collaborative R&D with local districts. These efforts have become standard operations for SWRL, so much so that we take the relationship for granted. However, the political, economic, professional, and technical aspects of a successful collaborative project are unique. Consequently, such endeavors are more talked about than tried. Furthermore, others who attempt them seldom succeed.

Conditions that are conducive to producing consensus, a prerequisite for collaboration with a district, are rarely present. For example, some product or procedure exists already; although it may not be working well, it is in place--with people's history and future attached to it. Additionally, the situation usually includes a department or staff that does not see the need for outside help, one that has invested long hours and personal ego in setting some things into place. Moreover, the outside agency is usually dealing with the department indirectly through a district administrator.

Since informed consensus is not present at the beginning of the project, it must be earned and maintained. Establishing that status takes work--characterized by honesty, integrity, and negotiated reciprocity. The work is not aptly characterized by a familiar designation such as research, development, evaluation, or dissemination. Rather, it is all of these, pieced together by negotiations comprised of sequential communications. These communications operationalize the momentum of the effort.

The sequential communication moves only so fast as the district and SWRL are willing and able to move. The effort proceeds as rapidly as SWRL can provide the district with whatever is required before SWRL can take further action. The moving effort inherently provides the district with regular opportunities to react, interact, alter, and initiate. The district and SWRL are mutually involved in changing and modifying only the effort, not with changing or modifying either the district or SWRL. Furthermore, the concern is not to do all the work "together," but to take advantage of SWRL's and the district's respective and complementary strengths and capabilities.

SWRL's intent is to establish a situation that is consistent with the state of the art technically, while achieving a positive consensus among district participants. This intent creates, by the end of the project, a situation that is both professionally sound and operationally institutionalized within the district.

This paper describes such an activity, a recent project in which SWRL designed and developed a minimum standards test. The paper first presents background information, and next discusses, in separate sections, the participants, materials, and procedures involved in the activity. It then details the results of the activity and concludes with a discussion of compromises made by both SWRL and the district. "Between the lines," the paper describes the procedures involved in collaborative projects.

BACKGROUND

The district involved in the project is a large urban school district in California. The activity described focuses on the writing

sample included in the district's minimum-standards test for high school students. The original prompt asked students to write something that they remembered about their impressions on their first day at school.

A review revealed that although the District did provide a topic that would be part of every student's experience, the prompt could pose problems for students who might not remember their first day at school; they would have to make up an experience. Thus the topic would increase the testing burden for those students, who would also be assessed on their creative ability, while students who remembered a first day at school would not.

Another problematic aspect of the prompt was noted in the review: It would assess skills that are disparate in kind and number across students, for students would likely interpret the purpose of the task differently. Specifically, one statement in the prompt asked for "impressions," so some students might write descriptive essays; other students, focusing on the time feature in the last sentence of the prompt (i.e., "Think of the day from early morning to the time you arrived home") might write personal narratives. These compositions differ in kinds of linguistic structures, diction, and organization, as well as in type of content, so writers must use different skills to compose them. For example, "sensory impressions" may prompt spatial ordering in descriptions, while significant events may prompt chronological ordering in narratives. Furthermore, the presence of different sets of features in descriptive and narrative compositions correlates with holistic ratings for those compositions (Humes, 1983).

Because of these complications, and because other types of discourse are more important than personal narratives for functioning outside school, SWRL confirmed the district's practice of drawing on the experiences of all students, but suggested that a new prompt be devised that did not depend upon students' memories of an event that occurred possibly 11-12 years earlier. We suggested that the new prompt assess the same skills for all students, particularly their ability to communicate information clearly and persuasively, a proficiency important for success both in college and in the workplace. To move quickly past likely initial concerns about this change in purpose, SWRL recommended that the prompt be tried with a sample of district students before administering it in an actual minimum standards test. This approach shifted the immediate attention and activity of district participants into high gear in order to set up and obtain the feasibility experience. All necessary logistics were left to administrators and school-site personnel. By fiat, the activity moved into another stage, leaving the door open to return to the old format.

PARTICIPANTS

Three readers were selected by the school district to read and evaluate students' writing samples. These readers had been trained to use the district's holistic scoring procedures and were judged by the district to be among its best evaluators. This selection was important, because it meant we had three participants who were very likely to give the new alternative a good opportunity to work.

Student participants, selected by district personnel, were thirty-five high school seniors enrolled in summer school classes in order to meet school requirements for promotion. The group was comprised of Chinese, Vietnamese, and Hispanic students whose first language is not English. This student sample was not optimally representative, but it introduced a healthy conservative bias to the inquiry: If an instrument administered at twelfth grade errs conservatively, it will err by passing--not failing--a student. We knew that if we had consensus among three top raters and "passed" about 50% of these students, then our instrument would be both sound and conservative.

MATERIALS

SWRL developed both the assessment and the scoring materials. Assessment materials included 35 copies of the new writing prompt prenumbered in red from 1 to 35, 35 student response sheets also prenumbered in red from 1 to 35, and several copies of directions for administering the new prompt.

The prompt was designed to elicit written discourse that communicates information clearly and persuasively; it asked students to write a letter to the fictitious manager of a nearby restaurant. Students were to convince the manager that they should be hired for a specific job by naming that job, stating reasons for the hiring, and giving support for those reasons. Letter-format skills were not assessed by this prompt because the heading, address, and greeting were printed on the students' response sheet. However, the response sheet did not have a

printed complimentary closing because its appearance would prescribe the length of the letter. Furthermore, the response sheet consisted of two pages for writing so that students could compose a long letter if they wished.

Scoring materials developed for the pilot study included directions that explained how to score and record the ratings, for which a Record Sheet was also provided. Guidelines for determining the ratings were also part of the scoring materials. The guidelines explained how to rate the samples by each of four different methods (see Procedures section for a discussion of these methods; see the Appendix for the Directions and Scoring Guidelines). These materials were packaged into three sets labeled "Scorer 1," "Scorer 2," and Scorer 3."

Using four scoring procedures is unconventional, but we suggested the procedures for two operational reasons and two background ones. Operationally, multiple methods produce (1) different kinds of information about evaluators' consistency in rating the samples, and (2) evidence of the usability of the prompt, given students' success under different scoring conditions.

The background reasons reflect general methodological concerns about scoring compositions. SWRL often recommends a descriptive approach to scoring, which was appropriate here. Adding two scoring procedures in addition to our descriptive and the district's holistic approaches allowed us to explore these methods while obviating any interpretation of "our method vs. your method." Holistic scoring is, by design, a relative process, the "standard" depending on how samples fall into stacks, as

well as on how well trained the scorers are (Humes, 1980). Furthermore, the validity of the method as a measure of writing competence is increasingly being questioned (Kinneavy, 1980). Holistic assessment has been popular with school districts because the National Writing Projects that promulgate this method for use in the classroom are influential and because the prompts and the scoring are easier to design: the prompt need not elicit every feature on which all students can be scored, and, therefore, specific scoring criteria need not be developed.

PROCEDURES

First the prompt described above was designed, reviewed, and revised, using conventional SWRL procedures (see Humes, 1980). Then the scoring materials were developed. These materials were mailed to the district for use in conducting the pilot study.

Teachers distributed materials to students participating in the study, giving them ten minutes to read the prompt and plan their compositions. They were then allowed thirty minutes for writing. At the end of that period, teachers collected both the writing prompts and the response sheets from the students.

These materials were returned to district personnel, who collected and organized the materials for scoring, dividing them into four relatively equal stacks and labeling them with the letter "S," for "scoring method," and a number. All compositions in the first stack were marked with the code "S1" for scoring method 1, in the second stack with S2, in the third with S3, and in the fourth with S4. These four stacks

were then distributed to the evaluators participating in the study. Each evaluator scored all four stacks, using the method noted for the stack--S1, S2, S3, or S4. As they scored a sample from a stack, evaluators noted on the Record Sheet the student's response-sheet number and his or her rating on the writing sample (see the Appendix for this document).

To score Stack 1, evaluators used the method in which district personnel were trained, true holistic scoring (as opposed to the modified versions noted below). As noted earlier, holistic scoring is a relative process: Papers are rated in relation to other papers in the set. For this study, evaluators worked with a scale having a range of four ratings (i.e., 1-4) because it is the range most frequently employed in holistic scoring to discriminate among writing samples.

To score Stack 2, evaluators used a method conforming to the SWRL model (Humes, 1980), in which readers follow a set of scoring guidelines developed specifically for a prompt. Guidelines developed for this method can be used without formal training, as is standard practice when this method is used by classroom teachers. Evaluators scored the samples on each of seven criteria that cover features characteristic of discourse that is elicited by this kind of prompt, i.e., reasons, proof for reasons, ordering, language, sentence structure, capitalization and punctuation, and spelling (see the Appendix). The scores assigned by the raters when they used this method depended on absolute standards described in the guidelines rather than on a rating in relation to other students' performances. The sum of the scores on the individual criteria determined a single, overall evaluation comparable to the single rating given when the scorers employed holistic methods.

Readers rated Stack 3 by using a criterion for standardizing content scores as a complement to holistic scoring, the method by which other features of the sample (e.g., organization, mechanics, spelling, language) were rated. This standardization criterion specified scores for numbers of reasons and their proof (see the Appendix). Scorers rated the sample on a 0-4 holistic scale, with a "0" rating to be given only when the student either did not write or did not write the task described in the prompt.

Evaluators scored the fourth stack by using a hybrid form of holistic and analytic scoring on a 0-4 scale. This uncommon method involved standard scores, as for Method 2, rather than relative ratings, as for Methods 1 and 3. Guidelines specified a variety of possible combinations of characteristics on which to evaluate the sample (see the Appendix). In using this method, evaluators had to distinguish among one description for a 4 rating, three descriptions for a possible 3 rating, two descriptions for a possible 2 rating, and one description for a 1 rating. A rating of "0" was given only if the student did not write or did not write the task in the prompt.

Completed materials were collected by District personnel and mailed back to SWRL for analysis. When the materials arrived, Method 2 samples were prepared for scoring. The good-acceptable-unacceptable ratings were converted to numbers, i.e., 3-2-0, for each of the seven component criteria. Next the scores on all criteria for each writing sample were added and then divided by the number of criteria. This procedure reduced the ratings to one overall score for each sample so that single scores

could be compared across the three evaluators and so that consistency could be compared across scoring procedures. Tables were then prepared to analyze and compare evaluators' performance when using each of the four scoring methods.

Finally, the writing samples were read to identify any problems students may have had in responding to the prompt and to insure that the task was appropriate for the population assessed.

RESULTS

Consistency of Scores

The analysis of the scoring revealed that scores were most consistent across evaluators when they used Method 2, the descriptive scoring procedure, despite evaluators' training in Method 1. This consistency is evident in the table displaying the correlations for the scorers across methods, which is presented at the end of this section. The results for each method are discussed below, with a table displaying the results complementing the discussion.

Method 1: Holistic

As previously noted, samples scored by Method 1, conventional holistic, were given one numerical rating by each scorer; these ratings ranged from 1-4. Table 1 reveals that, with two exceptions, Scorer 2 rated the samples one point lower than did Scorers 1 and 3. For those two exceptions, all three scorers gave the same rating.

In only one case did the scores range across more than two rating categories. The sample receiving these various ratings was read closely to determine if there were any reason for the discrepancy. This reading

Table 1

Method 1 (Holistic) Scores

Sample No.	Scorer 1	Scorer 2	Scorer 3	Total Score
3	X	2	3	5
7	1	1	1	3
8	2	1	2	5
16	3	1	2	6
27	2	1	2	5
28	3	2	3	8
29	2	1	2	5
30	2	2	2	6

suggested that Scorer 2 may have placed more emphasis than did the other scorers on correct form over appropriate content, and the student had a number of problems with the form skills of standard written English.

One sample was not evaluated by Scorer 1 (see the "X" in Table 1). However, the ratings recorded by Scorers 2 and 3 indicate that the writer of that sample received an acceptable rating according to the criterion for a passing score, i.e., a rating of 2 (one-half the highest possible score of 4), times the number of scorers. Thus four out of eight students received passing or better scores. Furthermore, three of the other four were just below the borderline because of the generally lower ratings of the one evaluator. This success rate is quite good when the language difficulties of the sampled students are considered.

Method 2: Descriptive

Table 2

Method 2 (Descriptive) Scores

Sample No.	Scorer 1	Scorer 2	Scorer 3	Total Score
1	3	2.6	2.9	8.5
2	3	2.6	2	7.6
6	.9	1.7	1.1	3.7
14	0	0	.9	.9
18	2.1	2.6	2	6.7
21	2	2.3	1.4	5.7
31	2.3	2.1	2	6.4
32	1	1.7	1.1	3.8

This method (Humes, 1980) produced the narrowest range of scores across raters (see Table 5 on page 17). In only one case did the scores for a sample range across as much as one full point. All other samples received scores with less spread across all three evaluators. These results are particularly encouraging since a single student received seven scores, each of which may have been as high as 3 or as low as 0, from each evaluator. (The method for reducing the seven criteria to one total score is explained above.)

Since the highest score on this scale is 3 when the good-acceptable-unacceptable ratings are converted to 3-2-0, a passing score per

evaluator must be 1.5, half the total. This score is thus comparable to the passing score for holistic scoring, described above, which is also half the total (i.e., 2 of a possible 4). The overall passing score for this method is also derived by multiplying by the number of raters, for a total of 4.5. Thus five out of eight samples received passing scores. The borderline sample suffered from uniformly poor scores on form criteria.

Method 3: Standardization Criterion

This method resulted in the least consistent scoring of the four methods employed for the pilot study (see Table 5). As noted above, this method complements holistic scoring with a standardization criterion for content. However, only one numerical score is given by a rater, using a scale with a 0-4 range.

Nine papers were evaluated using Method 3. One sample received the same rating from all three scorers, and six received ratings that ranged across only two categories. However, one sample received ratings that ranged across four categories (since 0 is a rating category in this system), and one sample received scores across three rating categories. These latter samples were read closely to identify any obvious reasons for this discrepancy. For one sample, good form skill usage seemed to have influenced two of the raters to give scores higher than the scores given by the other evaluator. For the other sample, the 0 score was given because the student did not write the task described in the prompt; rather, he wrote a letter about a non-restaurant job.

Five of the nine papers evaluated by this method earned passing scores (i.e., 2, times the number of raters). Three of the four that did not receive passing scores were on the borderline, with only one of the three evaluators giving the sample an unacceptable rating.

Table 3

Method 3 Scores

Sample No.	Scorer 1	Scorer 2	Scorer 3	Total Score
9	1	1	2	4
12	2	0	3	5
15	2	2	2	6
19	3	2	3	8
20	4	3	3	10
24	1	2	2	5
26	1	3	2	6
33	2	1	2	5
34	2	2	3	7

Method 4: Hybrid Holistic

Table 4

Method 4 Scores

Sample No.	Scorer 1	Scorer 2	Scorer 3	Total Score
4	4	3	3	10
5	3	3	2	8
10	3	3	3	9
11	3	3	0	6
13	2	1	2	5
17	2	1	0	3
22	2	0	0	2
23	2	1	2	4
25	3	3	3	9
35	2	2	2	6

This method resulted in less consistent scoring than did either Method 1 or 2, probably because the method required the evaluator to give a single score based on a number of co-occurring conditions for one rating, as described in the scoring guidelines (e.g., "Includes two reasons with proof for one or more reasons. Orders reasons coherently in generally acceptable language. Makes a number of errors"), and because evaluators also dealt with more than one such description for some scoring categories: three descriptions of possible 3 ratings, and two

descriptions of possible 2 ratings. In addition to these descriptions, the evaluators dealt with a description for each of the other possible ratings--4, 1, and 0. It is difficult to hold so many conditions and descriptions in short term memory while reading and evaluating a sample.

Despite this complexity, three of the ten samples received the same score from all three evaluators, with six of the ten samples receiving passing scores and four receiving scores that ranged across only two categories. However, one received scores ranging across four categories, and one across three categories. In fact, the range of ratings for two samples is strange because scoring categories are skipped (e.g., 3-3-0; 0-2-0). These two samples were read closely to determine reasons for the discrepancy in ratings. In one case, the student again wrote about a non-restaurant job, earning a 0 rating from two raters. No reason was identified for the discrepancy on the other sample.

Correlations

Pearson product-moment correlations were computed to determine consistency between scorers for the different methods. The correlations are presented below in Table 5.

The highest correlation was found between Scorer 1 and Scorer 2 for Method 2. The lowest correlation was found between Scorer 2 and Scorer 3 for Method 3. Other previously discussed consistencies and inconsistencies are evident from Table 5.

Table 5

Correlations between Scorers for Different Methods

Scoring Method	Scorers 1 & 2	Scorers 2 & 3	Scorers 1 & 3
Method 1	.35	.70	.84
Method 2	.91	.72	.87
Method 3	.26	.03	.71
Method 4	.82	.53	.48

Appropriateness of the Prompt

As previously noted, the writing samples were composed by students whose first language is not English. Most of these students had considerable difficulty with English diction and with the form skills of standard written English, i.e., syntactic structures, spelling (particularly affixation), capitalization, and punctuation. Despite their language problems, a majority of the students received acceptable or better scores. Students' samples were substantial in length, although no one wrote on the second page of the response sheet.

If scores had been based on content alone, nearly all students would have passed. Content in the samples indicated that students understood

the task and could respond to the prompt. Most students were able to write the three reasons requested in the prompt after stating the purpose of the letter (i.e., to apply for a particular job). Some students, however, gave personal rather than employer-oriented reasons why they should be hired. The prompt itself may have been responsible; it did not control for this kind of response since the wording did not specifically limit, other than by example, the kind of reasons the student could include. Some students had difficulty presenting reasons in an order of importance, but relative importance is determined somewhat subjectively.

These results suggested that the prompt would be appropriate after minor revisions. See the "Discussion" section below for a description of these changes.

DISCUSSION

The analysis of scoring revealed that Method 2 (descriptive) elicited the most consistent performance across evaluators, and "passed" about 60% of the students. However, there were no real borderline samples that would likely be moved into the "pass" category by making adjustments in the prompt and/or scoring procedures. Method 1 (holistic) produced high consistency among raters and "passed" about 50% of the students. Moreover, about 75% of the "non-passes" were just below the borderline, and adjustments in the prompt along with one or two procedural recommendations would likely move some students from just below to just above the borderline. Methods 3 and 4 both produced dramatic inconsistencies among raters and were only included as credible expedients for a broader purpose.

Based on these findings, SWRL advised the district to use the district's usual holistic scoring techniques to evaluate the 1983-84 samples. We made this recommendation for a few reasons. First, district raters were well-trained in the district's scoring procedure. Also, we knew we could fine-tune the prompt to improve inter-rater consistency and to force several borderline cases to better focus their compositions. This fine-tuning, along with two procedural recommendations (all described below) would generally tighten the holistic scoring process used in the district. Our last reason was pragmatic--there is less need for the diagnostic information provided by the Descriptive Method 2 when high school seniors are being assessed for minimum competency and the diagnosis won't be used as the basis for long-range instructional planning.

Two other scoring recommendations were made. One was that any samples not earning passing scores from ratings that range over three or more categories should be reread by one or more additional evaluators. The other recommendation was that readers be instructed not to consider the format skills of business-letter writing when they rate the samples because the prompt was not designed to assess these skills.

Three changes were made in the prompt itself. Wording was changed to suggest more closely that the students give employer-oriented rather than personal reasons for being hired. However, the wording was not changed to be so explicit that it excluded personal reasons, since evaluators should be allowed to give some credit for personal reasons on this minimum standard test.

The prompt was also revised to delete the statement about ordering (i.e., "Give your three reasons in order from least to most important"). This statement was in the prompt to cover the ordering criterion for Method 2 scoring since, according to this method, students cannot be evaluated on such a specific feature if they have not been cued to produce it. This explicit direction is unnecessary when holistic scoring is used, since specific features are not evaluated. Furthermore, this change will make the task both simpler to write and easier to evaluate.

The third change was not made because of any problems in the prompt; it was made to accommodate the machine scoring system used by the district. The scoring is marked on a blocked out area at the top of the response sheet, so nothing else that might be picked up during scanning can be printed on the page. Thus the heading, address, and salutation could not be printed on the response sheet as they were for the pilot study. To adjust for this condition, the prompt had to be changed as well. Since the assessment objectives of the writing sample do not include letter-format skills, the prompt was revised to tell students not to include letter parts other than the salutation, which was added to the prompt for students to copy.

The results of the feasibility inquiry suggested that the writing sample, as revised, is an appropriate task for the students, considering the performance of the students on the earlier version of the prompt and the relatively consistent scoring of the prompt under different scoring conditions, even those conditions for which the raters had not been trained.

SWRL had operated on two tracks during this piloting period, giving the district what it wanted while moving toward the direction of improved assessment. Rather than criticize the vulnerable shortcomings of their writing-sample prompt, we suggested that we could make it tighter. SWRL moved quickly on the design and development track, while giving the district frequent, regular opportunities to stop and return to earlier design on the negotiating track. Thus the district opted for the new prompt, and SWRL added the recommended scoring adjustments. The result was a sound and conservative instrument that had moved closer to the district's assessment objectives.

The activity entailed work for both the district and SWRL, but the work proceeded smoothly and all personnel involved experienced professional satisfaction in the task and its outcome. Other facets of the collaboration pertaining to the high school test in other subject areas proceeded similarly. The end of this project marks the <u>beginning</u> of an amicable cooperative relationship between SWRL and the local school district. This new cooperative situation was further cemented by the district and SWRL when both agreed to pursue R&D in connection with the district's junior high school and elementary school assessments as the next step in the collaboration.

References

Humes, A. A method for evaluating writing samples (Technical Note No. 2-80/02). Los Alamitos, Calif.: SWRL Educational Research and Development, 1980.

Humes, A. Textual features in the descriptions and fictional narratives of young writers (Technical Note No. 4-83/01). Los Alamitos, Calif.: SWRL Educational Research and Development, 1983.

Kinneavy, J. L. Evaluating the effectiveness of college writing programs. In A. Humes, B. Cronnell, J. Lawlor, and L. Gentry (Eds.), Moving between Practice and Research in Writing. Los Alamitos, CA: SWRL Educational Research and Development, 1981.

Appendix

PILOT SCORING MATERIALS

DIRECTIONS FOR SCORING

SCORING AND RECORDING THE WRITING SAMPLE

General Directions

You should receive four separate stacks of student compositions from [illegible] of the District. Each stack should contain 5 to 10 written compositions. You are going to score the stacks by using a different scoring technique on each one. The four distinct scoring options are described below:

OPTION 1: DISTRICT HOLISTIC SCORING

You are trained in the District's holistic scoring technique. If you are asked to use Option 1, you should use the procedures defined by the District for holistic scoring.

OPTION 2: GUIDED SCORING ON SPECIFIC CRITERIA

A set of guidelines are provided and can be used by teachers without formal training. You should carefully follow the guidelines provided, scoring each criterion as **G** (Good), **A** (Acceptable), or **U** (Unacceptable).

OPTION 3: COMPLEMENT TO HOLISTIC

This option describes an uncommon procedure. It provides standardization for content scores, while other features (e.g., organization, mechanics, spelling, language) are measured holistically, using the District's usual procedures for discriminating among scores 4,3,2,1 with a rating of 2 or more (or the number of raters x 2) as a passing score.

OPTION 4: HYBRID HOLISTIC

This option provides a complete 0-4 scale with a rating of 2 or more (or the number of raters x 2) as a passing score. This hybrid and uncommon form attempts to describe the variety of possible responses, yet limits the range to a manageable number of possibilities.

Scoring and Recording

Detach the Record Sheet from the end of this package. You will record all of your scores on this sheet.

Start with the first stack. Your record sheet tells you to score this stack by using the District's holistic procedure. Follow the steps below to score the first composition in the stack:

1. Record the prestamped ID number (in the upper right-hand corner of the student's sheet) on the first line.

2. Use the procedures for this option to score the composition, and record your score on the line beside the ID number.

Score all remaining compositions in this stack using this scoring option and recording the scores. When you have scored the entire stack, return it to Mxxxxfixhxixx and she will give you another stack to score.

Score and record the second stack of students' compositions, using the second option shown on your Record Sheet. Before you start, carefully read the attached procedures for this option, and refer to the procedures as often as necessary. When you have finished the entire stack, give it to Mxxxxfixhxixx and she will give you a third stack to score.

Be sure the third stack is not one you have seen before. Score and record the third stack of students' compositions, using the third option shown on your Record Sheet. Again, carefully read the attached procedures for this option, and refer to the procedures as often as necessary. When you have finished the stack, give it to Mxxxxfixhxixx and she will give you a fourth stack to score.

Be sure the last stack is not one you have seen before. Score and record this stack of compositions using the third option shown on your Record Sheet. Again, carefully read the attached procedures for this option and refer to them as often as necessary. When you have finished this last stack, give it to Mxxxxfixhxixx along with these directions and your Record Sheet and she will return the materials to SWRL.

SCORING GUIDELINES

Scoring Option

OPTION ONE

Holistic Scoring.

OPTION TWO

Submits a scorable composition:

Unacceptable: The composition is not scorable for one of these reasons:

a. It is illegible.

b. It is not relevant to the prompt.

c. Student did not respond. If this criterion is marked unacceptable, do not score further.

Content:

1. Includes reasons.

Good: Includes three reasons.

Acceptable: Includes one or two reasons.

Unacceptable: Does not include reasons.

2. Includes proof.

Good: Includes proof for three reasons.

Acceptable: Includes proof for one or two reasons.

Unacceptable: Does not include proof.

3. Use logical ordering.

Good: Orders content logically, putting an important reason last.

Acceptable: Orders content coherently, although an important reason may not be last.

Unacceptable: Does not order content coherently.

4. Uses appropriate language (e.g., "good worker" rather than "super worker").

Good: Uses only language that is appropriate for a business letter.

Acceptable: Uses language that is generally acceptable, but may use a few colloquial or slang expressions.

Unacceptable: Uses slang and colloquial language consistently.

Form:

5. Uses correct sentence structure.

Good: Does not use fragments or run-on sentences.

Acceptable: Uses a few fragments and/or run-on sentences.

Unacceptable: Uses many fragments and/or run-ons.

6. Uses correct capitalization and punctuation.

Good: Makes few or no errors in capitalization and punctuation.

Acceptable: Makes several errors in capitalization and punctuation.

Unacceptable: Makes many errors in capitalization and punctuation.

7. Spells correctly.*

Good: Misspells few or no words.

Acceptable: Misspells several words.

Unacceptable: Misspells many words.

*If the student has misspelled the same word more than once, count it as one misspelling.

OPTION THREE

4: Includes three reasons and proof for each reason.

3: Includes three reasons but no proof, or includes two reasons with proof for one or two reasons.

2: Includes only one reason, or two reasons without proof.

1: Does not include reasons.

0: Does not write, or does not write about the prompt.

OPTION FOUR

4: Includes three reasons and proof for each reason. Includes proof after the reason it supports. Organizes logically, with most important reason last. Makes few errors in sentence structure, grammar, capitalization, punctuation, and spelling. Uses appropriate formal language.

3: Includes three reasons, presented with generally acceptable language in a coherent order. Makes a number of errors.

3: Includes two reasons with proof for one or more reasons. Orders reasons coherently in generally acceptable language. Makes a number of errors.

3: Includes two reasons but no proof. Orders reasons coherently in generally acceptable language. Makes very few errors.

2: Includes two reasons but no proof. Makes a number of errors.

2: Includes only one reason.

1: Does not include any reasons.

0: Does not write, or does not write about the prompt.

RECORD SHEET

SCORER NO. __

Score **one** stack of student compositions using Option 1, the District's Holistic method.

Option 1: **HOLISTIC SCORING**
(DISTRICT APPROVED METHOD)

Student Number	Rating (1,2,3,4)
____	____
____	____
____	____
____	____
____	____
____	____
____	____
____	____
____	____
____	____

Score a **second** stack of student compositions using Option 2, the criteria guidelines.

OPTION 2: GUIDELINES FOR SCORING SEVEN CRITERIA

Rating on Seven Criteria (c^1 through c^7)

G = Good A = Acceptable U = Unacceptable

(If the student submits an unscorable composition, leave the row blank.)

Student Number	c^1	c^2	c^3	c^4	c^5	c^6	c^7
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___

Score a **third** stack of student compositions using Option 3, the Complement to holistic scoring.

OPTION 3: COMPLEMENT TO HOLISTIC SCORING

Student Number	Rating (4,3,2,1,0)
___	___
___	___
___	___
___	___
___	___
___	___
___	___
___	___
___	___
___	___

Score a **fourth** stack of student compositions using Option 4, Hybrid holistic scoring.

OPTION 4: HYBRID FORM

Student Number	Rating (4,3,2,1,0)
____	____
____	____
____	____
____	____
____	____
____	____
____	____
____	____
____	____
____	____

When you finish scoring this stack, give the stack, the directions for scoring and administering the compositions, and this Record Sheet to Maxx Schuler.